DOCUMENT RESUME

ED 128 366 95 TM 005 488

AUTHOR	Silver, Evelyn Stern, Ed.
TITLE	Declining Test Scores: A Conference Report.
INSTITUTION	Johnson (Lawrence) and Associates, Inc., Washington, D.C.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Feb 76
CONTRACT	NIE-C-00-3-0060
NOTE	44p.; Proceedings from Conference on Declining Test Scores (Washington, D.C., June 19-21, 1975)
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	Academic Achievement; *College Entrance Examinations; *Conference Reports; Curriculum Development; *Low Achievement Factors; Predictive Validity; Research Needs; *Scores; Standardized Tests; Student Ability; Student Attitudes; Test Interpretation; Test Validity; *Trend Analysis
IDENTIFIERS	American College Test; Scholastic Aptitude Test; *Test Score Decline

ABSTRACT

In response to increasing evidence of score declines with no apparent agreement as to meaning or causes, the National Institute of Education (NIE) sponsored a Conference on Declining Test Scores in June of 1975. The objectives of the conference were to (1) clarify the evidence and estimate the extent of test score declines; (2) review evidence for the seriousness and meaningfulness of the problem and assess the value of research in this area; (3) explore areas of agreement and disagreement among experts as to possible causes of the declines; (4) formulate research guidelines for efficient and effective investigation into score trends and possible remedies; and (5) identify NIE's concern for and responsiveness to recent reports of score changes which could have important social implications. There did not appear to be consensus on the reasons for the decline even after the evidence for the various viewpoints had been presented and discussed. But there did appear to be consensus that further research could, at the very least, narrow the options and begin to assess the importance of the reported score changes. (Author/BW)

Declining Test Scores

A Conference Report

The National
Institute of
Education
U.S. Department of
Health, Education and Welfare
Washington, D.C. 20208

DECLINING TEST SCORES

A Conference Report

Edited by
Evelyn Stern Silver
Lawrence Johnson & Associates, Inc.

February 1976

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
David Mathews, *Secretary*
Virginia Trotter, *Assistant Secretary for Education*

NATIONAL INSTITUTE OF EDUCATION
Harold L. Hodgkinson, *Director*
Andrew Porter, *Acting Associate Director for Basic Skills*

CONTENTS

TABLES

FIGURES

EXECUTIVE SUMMARY

In response to increased attention and concern on the part of educators, legislators and the American public (see Chapter One), the National Institute of Education (NIE) sponsored a Conference on Declining Test Scores. Held on June 19-21, 1975 in Washington, D.C., the Conference was called to:

- clarify the problem of test score declines;
- review evidence for the seriousness of the problem and the value of research in this area;
- explore areas of agreement and disagreement among experts as to possible causes and treatments of the declines;
- formulate research guidelines for efficient and effective investigation into test score trends; and
- identify NIE's concern for, and responsiveness to, this important national problem.

Both the American College Testing Program (ACT) and the College Entrance Examination Board (CEEB) have reported consistent declines in average (mean) college entrance examination scores over the past decade. ACT scores have declined about one standard score over the last ten years. Recent data from the CEEB illustrate a dramatic drop in Scholastic Aptitude Test Scores--comparison of scores earned by 1975 high school graduates with 1974 scores shows a decline of ten points on the verbal part and eight points on the mathematical section.

The forty participants at the Conference on Declining Test Scores included representatives of test developers, test publishers, universities, private research centers and Federal agencies such as the NIE (see Chapter Two). Vigorous debate and dialogue gave participants the opportunity to review existing data and research (see Chapter Three), and to explore mutual areas of interest for future investigative efforts.

The three conference days included discussion groups on (1) Policy Issues and Impact, (2) Analysis of Causes, (3) Psychometric Scales and Data Requirements, and (4) Guidelines for Research and Remedies. Each group generated specific recommendations for research, ranging from analysis of test items and response patterns to surveys of test-taker attitudes and motivation (see Chapter Four). Several themes were stressed by participants:

- The score declines on college entrance examinations are important and worthy of further investigation.

- There are other trends in test scores that are similar to those reported for the college admissions tests but which involve other segments of the student population--these test trends should also be studied.

- There is no single cause of test score changes--research should focus on patterns of causal variables.

- Existing data, which are plentiful, should be fully exploited before expensive new data collection is undertaken.

There are important policy and planning implications for NIE from the Conference on Declining Test Scores (see Chapter Five). Existing data must be gathered together and sifted for descriptive information about trends in test scores. This review of research and related literature should be used to develop a comprehensive research framework within which secondary analyses of existing data and new research efforts can be planned. The coordination of investigations in this area will also insure comparability of data sources. The research framework that is developed should focus on the broader issue of test score variability, so that researchers are not tied into narrow explanations of score declines.

Such a comprehensive research framework would certainly include a longitudinal study of student attitudes about taking tests and testing. Score declines on college entrance examinations and on science achievement tests have recently been attributed to changes in student motivation. The lack of existing research on attitudinal factors makes it imperative that planning begin immediately so that data will be available for future analyses.

Several areas of research recommended by conference participants included a breakdown of college entrance test score trends by skill area tested. Such analysis would serve to pinpoint the locus of score declines and, perhaps, help to focus on the causes of declines. The Admissions Testing Program of the CEEB has analyzed scores for reading comprehension, vocabulary and "written English" for the 1974-75 population of test-takers. Follow-ups on this breakdown should yield the type of specific trend data that are needed.

Similarly, further analysis of score trends by sex seems warranted. The American College Testing Program has noted that the ACT Composite has dropped one standard score for men and 1.6 standard scores for women. The nature and extent of sex differences deserve further study so that causal factors may be isolated.

Finally, participants urged that NIE serve as a clearinghouse for research on test score trends. This function would insure coordination of diverse research efforts and accessibility to the data necessary for continued research and status reports.

The Conference on Declining Test Scores initiated an overview and investigation of this important topic. Speculation about the causes of score declines without empirical or research basis can lead to drastic and incorrect decisions--decisions that may affect educational personnel, programs, and priorities. Before remedial or preventive action is advocated, we must thoroughly investigate the evidence for score declines, analyze the possible meanings associated with such score changes, and then decide from educational research what factors might be responsible. Further research appears necessary in view of the diversity of opinion in these matters.

CHAPTER 1

THE NEED FOR A CONFERENCE

On September 7, 1975, the College Entrance Examination Board (CEEB) issued a press release highlighting the major findings of their report on college-bound seniors in 1974-75. The story appeared on the front pages of many Sunday newspapers (including The New York Times and The Washington Post) and revealed that average scores on the Scholastic Aptitude Test (SAT) had declined ten points on the verbal section and eight points on the mathematical section over the preceding year. The 1975 data followed a general pattern set by the past ten years of a downward trend of SAT scores.

A similar decline was reported by the American College Testing (ACT) Program on three out of four test batteries (English, mathematics, and social studies). Scores on the ACT natural sciences battery remained essentially constant over the same time period. However, average ACT composite scores for college-bound students declined by approximately one standard score, or about one-fifth of a standard deviation, over the past decade. On a scale of one to 36, the mean ACT composite score has dropped from 19.9 in 1964-65 to 18.6 in 1974-75 (on the first four testing dates in that year). English scores have declined by about one standard score, mathematics scores have declined by about 1.5 standard scores, and social studies scores have declined by about 2.5 standard scores. The American College Testing Program has reported that, "because of the large number of students tested each year, [850,000 high school seniors in 1974-75], the observed trends are clearly not due to random fluctuations in test scores but rather reflect actual changes" (Ferguson and Maxey, 1975, p. 5).

The CEEB is equally concerned about the meaning of the changes in SAT scores. The Board has examined score averages over the past twenty years and notes that scores were stable from 1956 through 1963. In 1964, average scores began to decline, particularly on the verbal section of the SAT. After leveling off in 1968, declines became more acute until 1974. In 1975, scores declined dramatically in the largest one-year drop yet recorded.

On a scale of 200-800, the average verbal SAT scores have dropped 39 points (from a mean score of 473 in 1956-57

to 434 in 1974-75). Average math scores declined 24 points (496 in 1956-57 to 472 in 1974-75). Over this time period, the deviations from the yearly mean of verbal scores have remained fairly stable, but the deviations around math score means have increased (McCandless, 1975).

ACT and SAT score averages for college-bound seniors from 1966-67 to 1974-75 are listed in Table 1. The data are illustrated in Figures 1 and 2.

Media coverage of the decline in test scores has intensified over the past twelve months. Editorials and news articles on this subject have appeared across the nation in papers such as The Arizona Republic (Phoenix, Arizona), the News-Star (Monroe, Louisiana) and the Mail (Charleston, West Virginia). Media coverage reflects the growing public concern over poor student performance, a concern matched by increasing professional attention. Education U.S.A., Education Daily, the Bulletin of the Education Commission of the States, The Independent School Bulletin, and the Chronicle of Higher Education are just a few of the professional publications that have pondered the test score decline. Entries in the Congressional Record in 1974 and 1975 demonstrate that concern among political leaders is also escalating.

As scores have declined, speculation about the causes and meaning of the reported changes has increased. There are data available to investigate some of these changes, but because these data are scattered and not generally accessible to the public, there has been little empirical research to support the explanations receiving the most publicity. Some of these explanations are: the failure of the public schools to teach basic computational and verbal skills; the decreasing benefits generated by a costly public school system; and the change in the composition of the student body that is seeking college admission. These are not the only speculations--a bewildering variety of specific causes for score declines among college-bound students has been considered. Hypotheses, which may often reflect the interests of their supporters, include changes in college admissions requirements and decreasing reliance on test scores, changes in the attitudes of students about tests in general, sociological and cultural trends that involve competitive and selective pressures, increases in curricular diversity, and changes in the content of test items.

Under its legislative mandate to improve education in the United States through research, the NIE has followed such

Table 1.--ACT and SAT Score Averages for College-Bound Seniors 1966-67 - 1974-75

	ACT Composite	SAT Verbal			SAT Mathematical		
	Total	Male	Female	Total	Male	Female	Total
1966-67	19.4	463	468	466	514	467	492
1967-68	19.0	464	466	466	512	470	492
1968-69	19.4	459	466	463	513	470	493
1969-70	19.5	459	461	460	509	465	488
1970-71	18.9	454	457	455	507	466	488
1971-72	18.8	454	452	453	505	461	484
1972-73	18.9	446	443	445	502	460	481
1973-74	18.7	447	442	444	501	459	480
1974-75	18.3	437	431	434	495	449	472

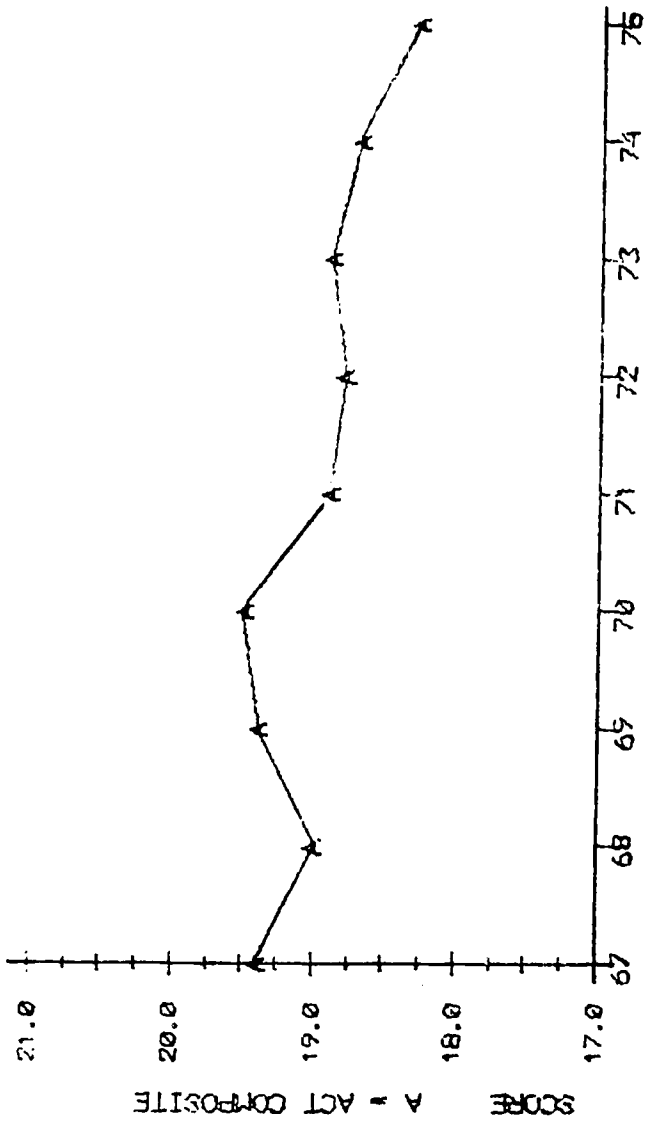


Figure 1.--*ACT Composite Score Trend*

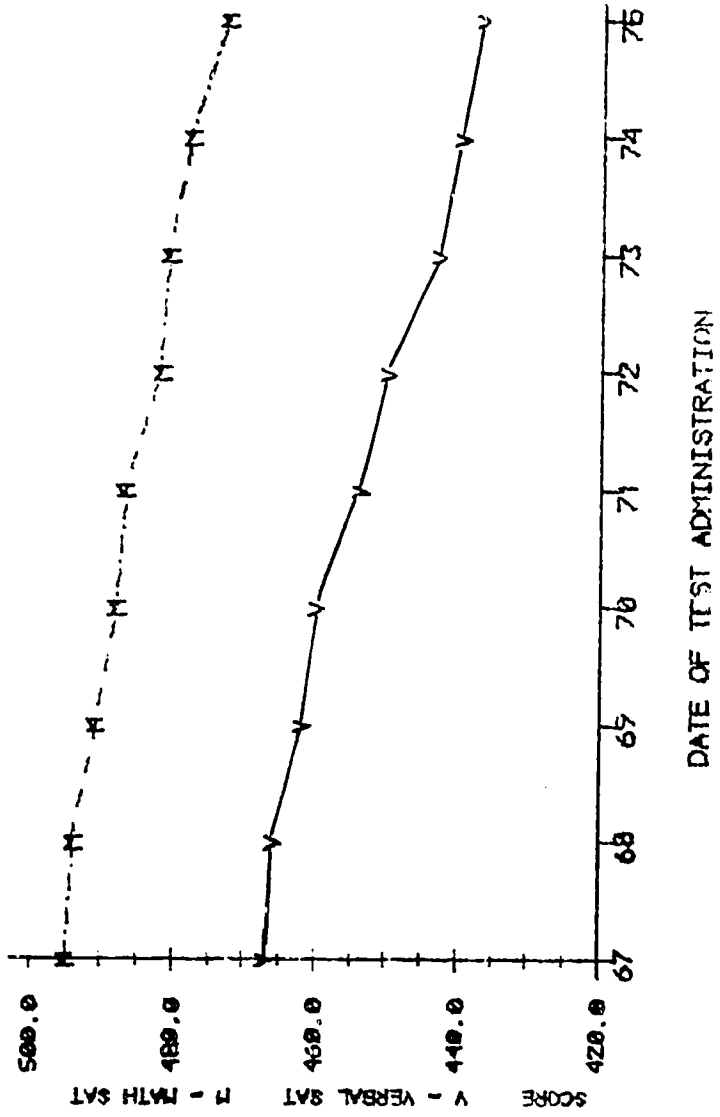


Figure 2.--*SAT Verbal and Math Score Trends*

speculation with interest. Because of the diversity of the speculation, it was thought appropriate to bring together many viewpoints to see whether a consensus would emerge or whether further investigation was necessary. The Conference on Declining Test Scores convened in June, 1975 was the beginning of this investigation. This report summarizes the viewpoints expressed and research proposed.

CHAPTER 2

CONFERENCE OBJECTIVES AND OVERVIEW

In response to increasing evidence of score declines with no apparent agreement as to meaning or causes, NIE sponsored a Conference on Declining Test Scores in June of 1975. The objectives of the conference were to:

- clarify the evidence and estimate the extent of test score declines;
- review evidence for the seriousness and meaningfulness of the problem and assess the value of research in this area;
- explore areas of agreement and disagreement among experts as to possible causes of the declines;
- formulate research guidelines for efficient and effective investigation into score trends and possible remedies; and
- identify NIE's concern for and responsiveness to recent reports of score changes which could have important social implications.

The Conference on Declining Test Scores was held from June 19 through June 21, 1975 in Washington, D.C. Forty participants included testing experts from universities, test designing and publishing organizations, research institutes, NIE and other Federal agencies. A list of attendees appears in Table 2.

The three-day schedule began with an opening address by the Director of NIE, Harold Hodgkinson, that set forth some of the evidence for the meaningfulness of score changes and the need for thoughtful investigation. Presentations followed concerning the nature of testing and test score interpretations as well as the public policy implications of alternative explanations for score declines.

Friday morning was devoted to short presentations followed by discussion on four problem areas related to score declines:

Table 2.--*Participants at the Conference on Declining Test Scores*

William Angoff	Educational Testing Service
Albert Beaton	Educational Testing Service
Elias Blake	Institute for Services to Education
Harold Bligh	Harcourt Brace Jovanovich, Inc.
John Carroll	University of North Carolina
James Cass	Saturday Review/World
T. Anne Cleary	College Entrance Examination Board
John Draper	McGraw-Hill/CTB
Roger Farr	Indiana University
Richard Ferguson	American College Testing Program
John Flanagan	American Institutes for Research
Walter Gillespie	National Science Foundation
Robert Guthrie	Office of Naval Research
Harold Harding	Yardstick Project
Lyle Jones	University of North Carolina
Hugh Lane	Institute for Services to Education
Sam McCandless	College Entrance Examination Board
Jason Millman	Cornell University
Amado Padilla	University of California at Los Angeles
Philip Rever	American College Testing Program
Don Searles	National Assessment of Educational Progress
Marion Shaycoft	American Institutes for Research
Robert Thorndike	Teachers College, Columbia University
Ralph Tyler	Center for Advanced Study in the Behavioral Sciences
David Wiley	University of Chicago
Dean Whitla	Harvard University
Belvin Williams	Educational Testing Service

From the National Institute of Education

Daniel Antonoplos	Garry McDaniels
Michael Cohen	Andrew Porter
Jane David	Jack Schwille
Linda Glendening	Marshall Smith
Harold Hodgkinson	Trevor Williams
Jackie Jenkins	Arthur Wise
Carlyle Maw	

- Summary of evidence for score declines--areas of declines, extent of declines, comparisons among tests and test-taking populations.

- Impact of declines--public opinion and response to news of declines; implications for public policy and conduct of research; questions of special concern to subject matter areas or to sub-populations of students.

- Causes of declines--historical and sociological correlates of declines; changes in tests, curricula, student populations and home environments.

- Remedies and Research--summary of possible interventions, research guidelines, further data requirements, and aid to schools and colleges for administration and planning; options for immediate action, long-term planning, and future preparedness.

In the afternoon, participants formed discussion groups to analyze causes of score declines, impact of declines on public policy, remedies and research guidelines, and psychometric explanations and data requirements for further research. Each group produced a summary of its discussion for review by all participants on Saturday morning. The conference concluded with a discussion of problem definition and recommendations for future research.

The vigorous discussions at the Conference on Declining Test Scores raised a number of important concerns and generated valuable ideas for research. There did not appear to be consensus on the reasons for the declines even after the evidence for the various viewpoints had been presented and discussed. But there did appear to be consensus that further research could, at the very least, narrow the options and begin to assess the importance of the reported score changes. Before summarizing the content of these sessions, Chapter 3 will review existing research to provide a background for the discussions and recommendations.

CHAPTER 3

RESEARCH ON SCORE DECLINES

An important outcome of the Conference on Declining Test Scores was the exchange of information on previous and current investigations into test score trends. A number of such investigations have been reported by the American College Testing Program (ACT), the College Entrance Examination Board (CEEB), and the Educational Testing Service (ETS). Other investigations of time trends, which had not been so well publicized or which had not been seen as relevant to the interpretation of score changes, were also discussed at the conference. In this chapter, four broad areas of existing research are reviewed:

- predictive validity of college entrance exams,
- stability of score scales,
- changes in test-taking populations, and
- changes in overall student ability.

The specific issues and research recommendations raised by conference participants (Chapter 4) are based upon the studies cited below and upon updated information about student populations and test scores.

Predictive Validity of College Entrance Examinations

The steady decline in test scores has raised questions about the validity of college entrance examinations. Since these tests have been designed to help forecast academic performance in college, test publishers have focused on considerations of predictive validity. It should be noted that more than test scores are required for adequate prediction; the producers of tests by and large agree that test scores should be used as supplemental information about college potential. This caveat is important to remember in understanding what the tests are designed to measure and what meaning can be attached to score changes.1/

1/ "The essential supplementary nature of the SAT is further attested to by the manner in which its effectiveness is

The predictive validity of the SAT is routinely measured by the Validity Study Service, through which CEEB provides colleges with analyses of how SAT scores are related to subsequent academic performance in college (McCandless, 1975). Usually, the freshman-year grade average is the criterion for academic success used in these validity studies.

Comparable studies for 1971 or 1972 freshmen and for three freshman classes since 1963 are available for 21 of the colleges that participate in the Validity Study Service. A comparison of these studies showed no evidence of a systematic increase or decrease in the predictive validity of the SAT (McCandless, 1975). Seventeen of the colleges had comparable data for 1967 and 1972 freshmen. For these colleges, the validity coefficient (a correlation coefficient) for the composite SAT scores (math and verbal) ranged from a low of .149 to a high of .602 in the 1967 studies, and from a low of .226 to a high of .632 in the 1972 studies. Since most of the SAT score decline is post-1967, it is interesting to note that the median validity coefficient for the 17 colleges was .421 for both time periods.

CEEB is cautious, however, about interpretation of this validity data:

> "This evidence that the [predictive] validity of the SAT has not been somehow affected by the score decline is drawn, of course from records that happen to be available for analysis after the fact and not from sets of data collected with stalking the SAT score decline in mind. But this is characteristic of the data that seem to bear on the score decline. All such data are fortuitously available and not exactly what we would like." (McCandless, 1975, p. 3)

ordinarily evaluated. Its simple (zero-order) correlation with college performance is by itself not a sufficient indicator of its usefulness in selection. Far more interest attaches to its incremental effectiveness, that is, to the degree to which it can improve the prediction of college grades when combined with high school records and Achievement Test scores. The SAT, therefore, is valued to the extent that it can add something unique to the other measures." (Angoff, 1971).

ACT has also found that the ability to predict college grades based on the ACT Assessment has remained stable. National norms from several hundred colleges that participate in ACT's Standard Research Service show that the multiple correlation of the ACT Assessment and first-year grade point averages has been very stable (about .5) over the 1966-1973 period (Ferguson and Maxey, 1975).

These ACT and CEEB findings are especially interesting in light of the steady increase in grade point averages at both the high school and college levels. ACT data, based on high school and college grades of students who attended approximately 400 colleges and universities which used the ACT predictive services from 1966-67 through 1972-73, provide evidence of this increase. Students at these colleges reported their latest high school grades prior to their senior year of high school for courses in each of four curriculum areas: English, mathematics, social studies and natural sciences. The mean of these four grades increased by .2 on a 4 point scale, over the six-year period. First semester college grade point averages of approximately the same group of students increased, on the average, by .27 units over the same time span (Ferguson and Maxey, 1975).

The increase in college and high school grades, known as "grade inflation," has received increasing public attention (Etzioni, 1975). The interaction of these two factors--increasing grade point averages and declining test scores--will probably result in closer scrutiny of the predictive validity measures of college entrance examinations.

Given the stated purposes of college entrance examinations, validity studies have focused almost exclusively on predictive validity. Yet several of the causal hypotheses for test score declines (e.g. changes in school curricula and in students' educational experiences) depend upon the specific content of the tests and the specific nature of skills assessed. The current focus on scores as indicators as well as predictors of student ability necessitates further research into other aspects of test validity. Work on content and construct validity is known to be difficult and has not been well researched in the past for a variety of reasons (Nunnally, 1975); nevertheless, further work in this area is indicated.

Stability of Score Scales

Several researchers at ETS have investigated the possibility that declining test scores result from the increasing difficulty of the tests. Every new form of the test is equated 2/ with earlier forms and scores are reported on a continuing score scale (Angoff, 1975). Over a number of years, however, small incremental changes in difficulty might result in a substantial "drift" in the score scales.

CEEB has long been concerned with the methods of equating test forms. The control and stabilization of test difficulty is achieved by retaining a group of calibrating or anchor items in several different test forms to allow comparison across forms used in different years. Of course this technique assumes that the group of unchanging items remains at a constant difficulty level in spite of changes in school curricula and course requirements.

Stewart (1966) investigated equality of SAT forms from 1944 to 1963 and Modu and Stern (1975) updated this investigation for the period of 1963 to 1973. In the latter study, two verbal and mathematical sections from 1963 and 1966 SAT forms were administered to two random samples of 1973 SAT test-takers. Modu and Stern concluded that the SAT may have been less difficult in 1973 than in 1963, and certainly not more difficult. Such a finding implies that the actual drop in ability levels may be greater than reported.

Studies of score stability, like the investigations of test validity, must be viewed cautiously for they attach special meanings to the words "stability" and "validity" that are different from common usage. It is difficult to avoid confounding score stability research with other variables such as test reliability and student ability. It is especially difficult to avoid such confounding in studies which range over a decade of history and which use non-experimental data. Again, it seems fair to conclude that basic research in this area has not received sufficient attention in the past and that such research should now be actively encouraged.

2/ Equating is a technical term for a number of techniques used to compare different forms of a test given to different groups of students at different times. For a summary, see Angoff, 1971.

Changes in the Population of Test-Takers

Some experts have accounted for the decline in test scores by showing that there have been changes in the characteristics of the population taking college admissions tests. These population changes may be caused either by (1) the inclusion of student groups who, in previous years, were not college-bound, or (2) changes in the characteristics or test-taking patterns of traditional college-bound students.

The first hypothesis assumes that the population of test-takers has expanded or become more diverse. According to this hypothesis, college-bound students are the highest ability students. As the proportion of high school students who go to college increases, the average ability level must drop. If this hypothesis is correct, test score declines indicate the success of egalitarian educational policies (e.g. Federal student loan programs, open enrollment and affirmative action policies, and special recruitment and counseling services).

Evidence of increasing test-taking rates is inconclusive, however. CEEB has examined the number of high school graduates, SAT-takers and college entrants as proportions of the total population of 18-year-olds, over time (McCandless, 1975). The rate of high school graduation for both sexes increased from 60-65% in 1959 to approximately 75% in 1965. Since then, the percentage of all youth graduating has stabilized at or near 75%. The proportion of high school students who go straight on to college has also increased since 1959. After a brief interruption around 1963, the growth trend continued until 1968-69, when it peaked at approximately 40% of 18-year-olds. Since then, the rate of immediate college entry after high school has declined five percentage points for females and ten percentage points for males.

Actual college admissions test-taking rates are difficult to obtain. Some indirect evidence indicates substantial increases, but only before the period of the observed score declines. In 1959, the number of students taking the SAT was about 50% of the number of immediate college entrants. By 1964, the number of SAT-takers was greater than 65% of the immediate college entrants; this percentage appears to have remained constant through 1973 (McCandless, 1975). Whether or not this constant percentage also implies an unchanging composition of student types and abilities deserves further investigation.

ge.

.s also of interest to note that the proportion of
:ing the ACT and the SAT has increased during the
ıde. In 1967, approximately 46% of SAT and ACT test-
:re women; by 1974, about 50% of SAT-takers and 52%
ıkers were women. Table 1 of this report reveals
ıle scores on the verbal section of the SAT have
ın average of 37 points from 1966-67 to 1974-75.
ıis same time period, male scores on this section
only a 26 point decline. Similarly, since 1965-66
:omposite has dropped one standard score for men and
lard scores for women.

:he proportion of women taking college admissions
increased, their scores have decreased more mark-
have the scores of men. Supporting evidence that
ıility and achievement trends deserve more research
comes from the National Assessment of Educational
(NAEP). Results from NAEP Assessments show that the
where females consistently outperform males is in
Males generally do better than females in mathe-
cience, social studies and citizenship (NAEP News-
ctober 1975). The NAEP results further show that in
.r subject areas, "males and females at age 9 show

. Recommendations:

scholastic understandings that are fairly equal. By age 13, however, females have begun a decline in achievement, which continues downward through age 17 and into adulthood."

Subgroup breakdowns by sex are only one way to analyze changes in the characteristics and test-taking patterns of the population taking the ACT and SAT. Investigators have examined other test-taking subgroups, including:

- college persisters: those students who remain in college after their freshman year;
- achievement test-takers: those students who not only take the SAT, but who also take one or more of the CEEB subject-matter achievement tests; and
- SAT repeaters: those students who take the SAT twice, in the junior and senior years of high school.

Rever and Kojaku (1975) analyzed ACT score trends among college persisters, those students who took the test in high school, enrolled in college and completed their freshman year. A sample of institutions that participate in annual predictive validity studies was taken from ACT's Research Service file. A comparison of the test scores of just those freshmen who subsequently completed their first year in college in 1966 and in 1973 yielded an important finding: no score decline. There were no uniform significant differences in test scores of students who completed their freshman year of college. Rever and Kojaku conclude that, "the data clearly suggest the tendency for lower-scoring students not to complete the first year of college" (p. 10).

The CEEB Achievement Tests are one-hour objective tests in fourteen academic subjects. About one-fourth of the students who take the SAT also take one or more CEEB Achievement Tests. Students take these tests to demonstrate subject matter preparation for college, to obtain placement in college level courses, and to fulfill admissions requirements.

Investigators have averaged the SAT scores over time of those students who take both the SAT and at least one CEEB Achievement Test. This subgroup of SAT-takers shows a markedly smaller decline in SAT scores than does the entire SAT-taking population. For some special groups of students who take specific CEEB Achievement Tests (such as physical

sciences or European history) the results show an increase in SAT scores from 1966 to 1974. The conclusion from this finding depends on whether this subgroup is a stable and definable part of the college-bound population. These results do offer some further indirect evidence that population changes can make a difference in score trends. This reinforces the need for detailed breakdowns of overall score trends in any thorough investigation of the reasons for score declines.

It should be noted that trends in CEEB Achievement Test scores have also been investigated, but they have not received as much public attention as the SAT scores. The CEEB Admissions Testing Program has examined score distributions for the seven most frequently chosen achievement tests as well as a score distribution for averages across achievement tests taken. The overall average score increased from 526 in 1971-72 to 531 in 1974-75. Several individual achievement tests showed slight declines over the same time period. In general, however, these data do not support the hypothesized decline in achievement for this group of test-takers. (College-Bound Seniors, 1974-75, 1975.)

Another area of research related to the hypothesis of a changing population of test-takers has been the study of the decreasing number of SAT repeaters (students who take the SAT exam as high school juniors and repeat it as seniors, either for practice or to better previous scores). The decreasing number of SAT repeaters may be due either to early admission in college for those who do especially well on the junior year administration of the test, or to more negative attitudes (decreasing motivation) for taking tests. Either reason could contribute to the score decline for the larger set of SAT test-takers. For example, if the most able students and highest junior year scorers were selectively less likely to take the senior year SAT because of early admission, then the decline in scores could reflect the fact that this group of able students was absent from the SAT-taking population.

McCandless (1975) found that although the number of repeaters has decreased recently, the number of such students is too small a proportion of the total SAT-taking population to have much of an effect (he estimates that the decrease in number of repeaters accounts for less than one-third of a point in the national mean in any single year).

Using a different methodology, Bullock and Stern analyzed SAT scores for 1966-67 and 1973-74 by removing all non-senior scores and by selecting high school students who took

the SAT on the same date. As a result of this grouping and the elimination of junior year scores, the mean decline of scores between these years was trimmed from 27 to 22 points.

Changes in Overall Student Ability

A hypothesis on score trends that has received national media attention relates score declines on college entrance exams to a more general decline in abilities of all students (not just the college-bound). This explanation has sometimes been accompanied by a criticism of educational programs and priorities. At the present time, however, research cannot link the score changes on tests for the college-bound to the probable score changes for the entire high school population because of sampling errors and non-random selection of students.

Other test data exist, however, which can be used to estimate trends in student performance over time. These tests are not college admissions tests and therefore are not subject to the criticism of a restricted population. There are other restrictions, however, on many of these scores (e.g. limitation to a particular state). Although these data are inconsistent and inconclusive, they are of interest both for what they indicate and for what they suggest for further research.

Flanagan and Jung (1970) administered a reading comprehension test in 1960 to a random sample of high school students (not just the college applicant population previously discussed) participating in Project TALENT. Ten years later the same test was given to a subsample of these same schools. Over this ten year period, the mean score in the test increased from 30.81 to 31.25.

Using data from the Preliminary Scholastic Aptitude Test (PSAT), administered in 1960, 1966, and 1970 in a random selection of schools, Schrader and Jackson (1975) found that mean verbal and mathematics scores had remained virtually constant. Although there may be sampling problems that could limit the generalizability of this finding, it is reasonable to conclude that the evidence does not support the hypothesis of a general decline in student abilities. More importantly, these data also show that the score decline does not exist for certain student populations even when SAT test items are used (the PSAT is composed of SAT test items).

NAEP data, referred to above, show an overall decline in student achievement. It is regrettable that the NAEP data collection effort was not begun sooner so that the time trends could be more easily compared, but there is across-time data available for three subject matter areas:

- science, 1969-1973;
- writing mechanics, 1969-1974; and
- basic reading/functional literacy, 1971-1974.

In science, there was a decrease in numbers of students who could answer most questions correctly. This finding applies to both 9 and 13 year-olds; the largest decrease in correct responses was for the older group. In writing mechanics an overall decline was noted for 13 and 17 year-olds, with a possible increase in correct responses for 9 year-olds. As before, it was the oldest group that performed most poorly over time.

In basic reading, there were some gains for the 17 year-olds. This finding was cited as a major inconsistency and used to challenge the evidence for a decline in other studies. Although it must be admitted that NAEP data (unlike other test data) are sampled specifically for nation-wide representation, this NAEP test battery measures very elementary skills, such as alphabetic ordering, reading road signs, and using telephone books. Accordingly, it may not reflect a very high level of intellectual functioning.

Some participants at the Conference on Declining Test Scores noted that it was in this area of basic literacy that national attention and financial support had been focused over the time period in which other score declines had occurred. For example, the Right to Read effort of the Office of Education and other compensatory programs have been concerned with functional literacy at the level tested by the NAEP instruments.

Another test that yields score data over time is the Minnesota Scholastic Aptitude Test, which has been administered to over 90% of all Minnesota high school juniors since 1959. The test yields a single score and is a measure of verbal aptitude. The scores on this test show an increase from 29.39 in 1967-68 to 34.71 in 1966-67. From that point on, a decrease in average scores is reported (31.05 in 1972-73). This is roughly the pattern of the ACT and SAT score trends for college-bound students.

The Iowa Tests of Educational Development (ITED) have been given to high school students in Iowa; results have been reported since 1962. The test includes seven skill areas, and is also reported in a single composite score. This composite score increased from 14.0 in 1962 to 14.5 in 1965, but declined to 13.5 in 1974.

The Iowa Tests of Basic Skills (ITBS) have also been part of the Iowa Testing Program; state-wide data have been reported for elementary schools since 1965. Average scores on most of the test batteries increased slightly or remained constant through 1966. Scores for the upper grades then declined until 1975. The lower primary grades, however, do not show a decline. Some conference participants noted that early elementary grades had been the major recipients of Federal and state support (e.g. Headstart, Follow-Through) throughout the period of other reported score declines and that such support may be related to the lack of a decline in the primary grades.

The Armed Forces Qualification Test (AFQT) contains four content areas: word knowledge, arithmetic reasoning, spatial perception, and knowledge of tool functions. AFQT scores from 1958 through 1972 are reported in Karpinos (1975). The results show that there has been continuous improvement in test scores throughout this period. The most significant change was the drop in the number of pre-inductees or draftees scoring below the tenth percentile score (a disqualifying score). Although the other percentile score groups therefore contain greater numbers of scores over the time period, the top ability group (the ninety-third percentile) actually shows a small decrease. This could be explained by changing deferment policies for college students during the 1960s.

CEEB has supported research on the hypothesis that changing student abilities account for the declines in scores. Using data from College Board "norm" studies in 1960, 1966, and 1974, the Board estimated SAT score averages for all eleventh graders (not just college-bound). The estimates showed increases in verbal ability between 1960 and 1966 which were then reversed between 1966 and 1974. Math ability reflected "trends" in the opposite direction; a decrease between 1960 and 1966 and an increase between 1966 and 1974 (The College Board News, 1975).

These estimates are projections for eleventh graders taking the SAT in October. The CEEB, based on other research, has hypothesized that SAT score decline may be partly

caused by a decrease in the growth of student abilities during the junior year of high school. Many students take the PSAT early in their junior year and then take the SAT in their senior year. A comparative study of seniors in 1967-68 and 1973-74 indicated that the amount of increase in test scores between the PSAT and the SAT testing was smaller for the 1974 class: the average gain in scores decreased 18 points (The College Board News, 1975). Once again, the reader should be cautioned concerning the acknowledged sampling restrictions.

This review of selected research on declining test scores illustrates the complexity of the problem which faced participants at the NIE Conference. In light of these facts and assumptions, participants raised the issues and recommendations included in the following section.

Readers interested in graphic and tabular displays of data discussed in this section are referred to the Technical Appendix of this report.

CHAPTER 4

CONCERNS AND RECOMMENDATIONS OF PARTICIPANTS

A summary of plenary and small group discussions at the Conference on Declining Test Scores is presented in this Chapter. The concerns raised at each conference session are followed by participants' recommendations for further investigation. Some of this material, and information in earlier chapters, was provided by participants during their review of the draft of this report.

Plenary Session

After an initial review of the major hypotheses of test score declines (see Chapter 3), participants focused on the inconsistent trends in ability tests. Table 3 lists examples of tests which have shown score declines and those which have not. In general, the tests for early primary grades do not show declines, while tests for upper primary grades and secondary school levels do.

Some participants suggested that the concentration of Federal funds which have created educational programs in the lower grades may account for the absence of test score declines in primary age groups. But it was also suggested that tests of this age group were concerned with "rote" learning and that the declines in later scores might be restricted to more abstract abilities.

Societal changes that could have contributed to downward score changes were discussed:

- categorical Federal aid for specific populations, rather than aid to the general school population;
- reactions to the Vietnam War and their effect on student moods, attitudes, and general disaffiliation;
- increased television viewing and other out-of-school activities which do not involve reading;

Table 3.--*Test Trends Over the Last Ten Years*

Declines	Increases or No Changes
American College Test (Composite)	Air Force Qualifications Test
Composite Test of Basic Skills	American College Test (Science)
Iowa Tests of Basic Skills (later grades)	Iowa Test of Basic Skills (early grades)
Iowa Tests of Educational Development	National Assessment of Educational Progress: Reading Achievement
Minnesota Scholastic Aptitude Test	Preliminary Scholastic Aptitude Test
National Assessment of Educational Progress: Science & Functional Literacy	Project TALENT
Scholastic Aptitude Test	

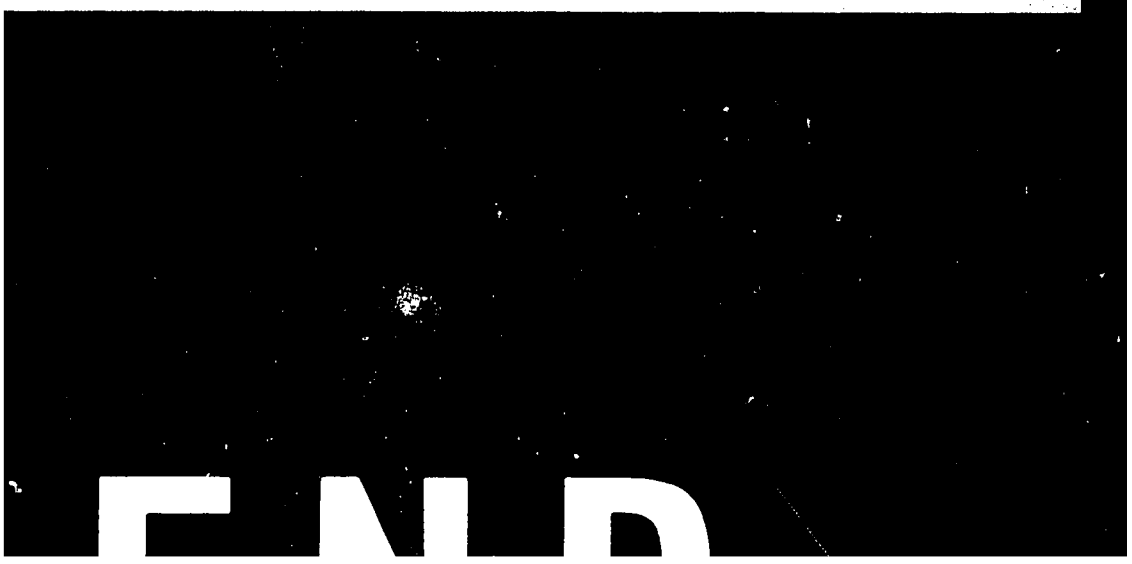
grades, retention in school, and graduation. On the
nd, both scores and college grades themselves have
w correlation with success in subsequent careers and
aningfulness was challenged.

day morning's plenary session set the stage for more
e small group discussions in the afternoon. Four
onducted simultaneous discussions on (1) policy is-
impact, (2) analysis of causes, (3) psychometric
nd data requirements, and (4) guidelines for research
dies.

oup Discussion: Policy Issues and Impact

ardless of the direction or magnitude of achievement
ends, participants agreed that public policy should
at providing students with a higher than current
performance with respect to verbal and mathematical
Public education offers substantial and continuing
ities for improvement and these opportunities should
ely supported.

areas of public policy were focused upon in this
(1) attrition of students during their first year
ge; and (2) cultural or social bias in testing. ACT
res have been declining for students applying to col-
llege-bound), yet the ACT scores of just those stu-
mpleting their freshman year have remained stable.
earch finding suggests that the score declines may be



associated with increased attrition in the first year of college, rather than with the quality of persisting college students. In other words, increased opportunity to enter higher education may not have resulted in increased opportunity to obtain a degree because of increased dropout rates in the first year of college.

The discussion of whether college entrance examinations fairly represent the ability to benefit from a college education included comments on the differential effect of timed tests on subgroups of test-takers and on alternatives to paper and pencil assessment. There was general agreement that tests were an improvement over the subjective and arbitrary selection procedures they had historically replaced. The history of the college admissions tests was cited to illustrate that these tests effectively opened the doors of elite colleges and universities to the middle class. But there remained questions of fairness to minority and lower socioeconomic groups.

Recommendations:

- Utilize existing data, e.g. National Assessment of Educational Progress, to reanalyze test data of comparable groups to corroborate college-bound score decline with different data bases.

- Survey teachers, school administrators and students to study their perspectives on the meaning and causes of score declines and of inconsistent score trends (non-declines).

- Evaluate relevance of specific test items used in equating protocols to give content validity and interpretation to score decline and to identify the "meaning" of the score decline with respect to item type.

Small Group Discussion: Analysis of Causes

This discussion group explored five hypotheses related to test score declines: changes over time in (1) the test-taking population, (2) school curricula, (3) attitudes toward taking tests, (4) test content, and (5) test-taking patterns. Each area of discussion is summarized below, with accompanying

research recommendations. General recommendations growing out of the afternoon discussion are also included.

In considering changes in the test-taking population, participants cited a need for data on the stability of selected high school populations and on subgroup analyses to show test score changes within these subgroups. Additional demograpnic data, by geographical area, were seen as useful. The participants also urged quantitative and qualitative assessment of sociological trends: attitudinal surveys, drug use patterns, number of single parent homes, number of working mothers, etc. Sociological trends should be studied in relation to specific test patterns. For example, patterns of television viewing may affect out-of-school reading patterns (a shift from reading to auditory comprehension). The same sociological changes may have different effects on the abilities of different age groups.

Recommendations:

- Analyze geographic and demographic breakdowns of available data on test scores over time.

- Analyze test score trends by type of institution (research universities, four year private schools, state colleges and universities, community colleges).

- Reanalyze ACT and SAT data by student subgroups to test hypotheses about changing populations as the cause of score declines on college admissions tests.

- Conduct correlational studies, i.e. television viewing and reading patterns over time in specific subgroups of students.

Participants recognized the increasing diversity of public school curricula. It has not been possible for test publishers to incorporate all of these changes in their tests to give full representation. Many publishers have instead tried to diminish the effect of specific curricula by avoiding all "method-specific" questions and by staying on ground common to all curricula. Questions remained, however, about whether test score changes could be related to the type of school curriculum offered and skills emphasized. The content and skills taught by curricula developed in the 1960s may not be fully

covered in tests which are equated to test forms developed in the 1940s and 1950s.1/

Recommendations:

- Identify aspects of school structure, school practice, and teaching activities that show change over time and relate such changes to test score trends.

- Analyze the possible effects of changes made in tests and item selection procedures which were designed to avoid dependence on particular curricula (tests were made less method-specific and designed to eliminate the advantage of having been taught a specific curriculum; also tests were "balanced" to include representative sections from different curricula).

Student attitudes toward test-taking may have shifted over the past decade. Participants believed that anxiety of test-takers may be an important variable to study. Similarly, the relationship between student motivation and the specific type of college to which the student aspires may be a fruitful area of research. Schools that select national student populations may show more or less decline in scores than do those institutions enrolling students from a restricted geographic area, and perhaps attaching different importance to test scores.

Recommendations:

- Conduct attitudinal surveys over time, secondary analyses of existing data, and retrospective surveys to isolate motivational changes with regard to test-taking.

- Collect data concerning test-taker anxiety (frequency of no-shows, noncompletions, attendance at review courses, and theft attempts, as possible proxy variables).

1/ The stimulus for criterion-referenced testing is in part a response to such issues. To the extent that nationally standardized tests focus on "general" educational objectives and exclude curricular changes at local levels or changes in local educational objectives, this concern with content validity seems justified.

In the process of test construction, especially with concern for the continuing relevance of entrance examinations in a rapidly changing society, the content of the tests may have changed significantly over time. In addition, the effort to broaden test relevance for new groups of test-takers may be adversely affecting the group for which the tests were originally developed (traditional college-bound students).

Recommendations:

- Analyze response patterns to test items which have reappeared on tests over a ten year period; reanalyze patterns for different population subgroups, if possible using information about curricular patterns over time.

- Conduct detailed analysis of specific skills tested by college entrance examinations and, where pertinent, of curriculum changes over the past decade. Examine score trends by skill (conceptualization, computation). These analyses should be set in the context of curriculum changes over time.

Aspects of shifting test-taking patterns that may impact on score changes include the time of year that tests have been administered and the proportions of students taking the tests on different testing dates. A gradual change in test dates over several years may account for some of the score changes, especially when coupled with admissions patterns for colleges (the best students can get early admission and need not retake the test).

Recommendation:

- Analyze score trends by test date and include separate analyses for students who re-take the tests, with information on college choice, college attended, and admissions policies (when such information is available for the time period of interest).

During the discussion, other ideas were introduced which did not fit easily under the above headings. These are listed below as general recommendations.

General Recommendations:

- Examine international data to help isolate and test cultural or societal factors (e.g. television viewing habits, curriculum changes, admission policy patterns, etc.).

- Study differences, and impact of differences, between ability and achievement tests in item design and item selection procedures.

- Compare different tests (content profiles, subject matter differences, etc.) to probe causes of inconsistent trends in test scores.

Small Group Discussion: Psychometric Scales and Data Requirements

Participants in this group agreed that test score changes are large enough in size and important enough in interpretation to warrant further investigation. Such investigation will be most cost effective if begun soon, since meaningful results require longitudinal studies. Also, uninformed speculation about score declines, which unhappily receives widespread media attention, may lead to incorrect decisions that could adversely affect educational programming.

Although it was generally doubted that the drop in test scores could be attributed to the mathematical techniques used to equate scores of different tests taken at different times by different groups of students, it was nevertheless felt by some that there was a need for detailed study of these procedures and better dissemination of the techniques so that they could receive discussion. Also identified was the need for more information on test-taking subgroups, details of score changes over time, and specific tests and measures showing changes.

The group cautioned planners that most of the research that can be done in this area will not be rigorously experimental and therefore will lead to correlational, not causal, information. Participants agreed that the general focus of future research should be on source of score variability over time, rather than on just the reported declines in college admission test scores.

Recommendations:

- Gather and systematically analyze existing data; focus on new variables for further hypothesis testing in existing data sets; investigate effect of grade inflation on validity coefficients for college admissions tests; investigate and compare procedures for equating test scores over time.

- Conduct comparative analysis of Project TALENT (1960), Coleman (mid 60s) and National Longitudinal Study (1972) data to corroborate college test score trends.

- Analyze PSAT, norms studies, state assessment tests, and follow-up studies which contain demographic and attitude information and which do not show score declines.

- Survey school systems to determine availability of test data and willingness to cooperate in data collection; data should include scores on tests, indicators of curriculum content, and relation of test scores to admission practices for college-bound students.

- Continue support of National Assessment of Educational Progress (NAEP); NAEP should (1) develop indicators of educational progress and (2) collect data on more variables that could lead to experimental studies.

- Examine tests at item and response levels; look at alternative scoring procedures to isolate content of decline scores.

- Review literature (including studies in press) to develop a matrix of existing data; agree on set of descriptors to insure comparability of planned research and establish a clearinghouse to coordinate and disseminate information.

- Conduct descriptive and comparative studies of ACT, PSAT, and SAT test-takers, and of college-bound and non college-bound students.

Small Group Discussion: Guidelines for Research and Remedies

Participants debated the meaning of test scores, questioning whether the scores were directly interpretable or were relevant only as proxies for specific educational achievements. The group agreed that a first step in further research on decline in scores was to define the important questions in terms of test validity (predictive, content and construct validity). Researchers should recognize that different constituent groups (parents, students, college admissions offices) will be interested in different questions and findings, and will have their own interpretations and emphases for test validity. Research should be general enough to be useful to this broad constituent group.

Once again, participants called for the assessment of student attitudes about tests. Within this area, researchers may want to focus on the effects of grade inflation or other admission requirements on motivation for test performance and on changes in the population of students taking the test (e.g. non-repeaters resulting from early admissions or open enrollment policies).

Recommendations:

- Conduct studies to define the areas of investigation that are important to different interest groups (parents, administrators, etc.).

- Study public policy questions related to skill levels, e.g. what level of reading is critical for performance of jobs essential to society? What level is required for college completion?

Concluding Session

Following a presentation by a representative of each small group session, the conference concluded with a general discussion of the topic and of future research directions. Three themes that had been mentioned throughout the meetings were reiterated: the decline in college entrance examination scores is large enough and consistent enough to be worthy of further investigation; we should not assume that there is a single cause or correlate of test score changes (thus the emphasis should be on patters of important variables); and

existing data, which are plentiful, should be fully exploited before expensive new data are collected.

Participants aired both agreements and disagreements as the conference drew to a close. They differed on whether shifts in the population of test-takers could be a significant contributor to score decline, and on whether the purpose of the ACT and SAT was to select (screen) college applicants or to assess student ability. Participants agreed, however, that the phenomenon of declining scores was both verifiable and statistically significant, and that the Federal government should guarantee support of a continuous monitoring system to avoid "panic research" and speculation.

Recommendations:

- Guarantee a systematic Federal research effort to assess the broad impact of score trends.
- Examine the possibility of different explanations for score trends in primary age groups and in college-bound groups.
- Systematically track national changes in primary and secondary school curricula; formulate educational indicators which can be relied upon for this purpose.

CHAPTER 5

POLICY AND PLANNING IMPLICATIONS

The persuasive evidence for a declining trend in scores on college entrance examinations coincides with increasing concern about test bias and educational accountability. The overlap in these three areas of attention has generated an unusual amount of public, professional and news media speculation. The Conference on Declining Test Scores concluded with a recommendation that NIE should assume a major responsibility for emphasizing objective and careful investigation of these problems and for coordinating the multi-faceted research into a comprehensive and useful framework.

Most of the diverse explanations offered for test score decline relate to the fundamental testing concerns of reliability and validity. Researchers in this area should be identifying and estimating the sources of variation (over time) both in individual test scores and in classroom or school district averages of scores. A possible research framework might include examination of four broad sources of variance in individual scores: family background characteristics, educational variables, individual psychological variables, and societal/cultural factors. These broad areas include the major variables discussed at the Conference on Declining Test Scores.

These four factors jointly influence high school performance and test scores, which in turn influence college selection, grades, and graduation. Finally, all of these factors may influence occupational satisfaction and success. The role of tests in this process and the interpretation of score changes with respect to this process need clarification and debate.

With regard to test validity, the conference discussions highlighted a lack of clarity in what the college entrance examinations measure and in whether such measures can be used as predictors equally well with different groups of students. Such tests are sensitive to individual differences in student ability, but there is some question as to whether they are also sensitive to other individual, educational, or societal influences. Although the predictive validity of these test scores for college grades has been demonstrated, the grades themselves do not appear to be reliable predictors of

post-schooling variables. The evidence of grade inflation seems to open suspicion about the stability of grades as a predicted or a predictor variable. Concern for validity is heightened by the lack of evidence that these test scores correlate well with factors other than college grades and persistence in college. Future research in the area of test validity should not focus exclusively on predictive validity coefficients and should use definitions of reliability and validity that reflect diverse social concerns.

The interest in score decline on college entrance exams has generated a great deal of activity in the research community. David Wiley, of the University of Chicago, is preparing an internal policy report for the Ford Foundation. The CEEB has announced the formation of a "blue ribbon" study panel. The Educational Research Service is studying declines in secondary level achievement test scores. Several conferences on this general topic are planned, including a seminar sponsored by the Edison Foundation and the I/D/E/A/ and a session on "Score Trends in National Exams and Their Implications," to be sponsored by the American Educational Research Association in April, 1976.

There is a significant role for NIE to play in the investigation of the declining test score phenomenon. As suggested by the conference participants, a first step is a comprehensive review of existing research as an aid in the development of a comprehensive research framework. NIE could serve as a clearinghouse and reduce duplication of effort while making data available for secondary analysis. Simultaneously, however, NIE should attempt to place the concern about test score trends in its proper perspective. To do so requires supporting research which will lead to a sound theoretical framework dealing with factors that affect test scores of both individual students and population subgroups. Only when such a framework is available can we relate test score changes to other educationally and socially important factors.

References

"A Million College-Bound Seniors Described in New ATP Report." The College Board News, September 1975.

Angoff, William H. "Scales, Norms and Equivalent Scores." In Educational Measurement, (2nd Edition). Edited by R. L. Thorndike. Washington, D. C.: American Council on Education, 1971.

Angoff, William H., ed. The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests. New York: College Entrance Examination Board, 1971.

Breland, Hunter M. A Summary of Research Relating to the SAT Score Decline. Princeton: Educational Testing Service, [1975].

Bullock, R., and Stern, J. Estimated SAT Summary Statistics for High School Cohorts 1967-1974. Princeton: Educational Testing Service, [1974].

College-Bound Seniors, 1974-75. Iowa City: Admissions Testing Program of the College Entrance Examination Board, [1975].

Etzioni, Amitai. "Grade Inflation." Science 190 (October 10, 1975).

Ferguson, Richard L., and Maxey, E. James. Trends in the Academic Performance of High School and College Students. Iowa City: The American College Testing Program, 1975.

Karpinos, Bernard. AFQT: Historical Data (1958-1972). Alexandria, Virginia: Human Resources Research Organization, [1975].

"Males Dominate in Educational Success." NAEP Newsletter, October 1975.

McCandless, Sam A. "The SAT Score Decline and Its Implications for College Admissions." Paper presented at the 1975 Western Regional meeting of the College Entrance Examination Board, January 1975.

Modu, Christopher C., and Stern, June. The Stability of the SAT Score Scale. Princeton: Educational Testing Service, [1975].

Nunnally, Jum C. "Psychometric Theory--25 Years Ago and Now." Educational Researcher 4, (November 1975): 7-21.

Rever, Phillip R., and Kojaku, Lawrence K. Access, Attrition, Test Scores and Grades of College Entrants and Persisters: 1965-1975. Iowa City: The American College Testing Program, [1975].

Stewart, Elizabeth E. The Stability of the SAT-Verbal Score Scale. Princeton: Educational Testing Service, [1966].

APPENDIX

The Technical Appendix summarizes data on the various tests discussed in this report, including date of testing, population tested, and other background information. Because these details may not be of general interest, this Appendix (and additional copies of this report) is available only upon request from:

National Institute of Education
Measurement and Methodology Division
1200 19th Street, N. W.
Washington, D. C. 20208

(202) 254-6512